Non-asymptotic, universal confidence sets for intrinsic means by mass concentration

Joint work with M. Glock, T.Hotz

Stefan Heyder
TU Ilmenau

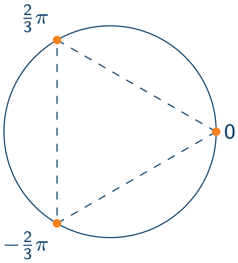
Fréchet means for circular data

- consider i.i.d. circular data $\theta, \theta_1, \ldots, \theta_n \in \mathbf{S}^1 \cong (-\pi, \pi]$
- equip $\mathbf{S}^1$ with arc-length metric d and recall the definition of the set of Fréchet means

$$M = \arg\min_{\mu \in \mathbf{S}^1} \mathbf{E} d^2(\theta, \mu) = \arg\min_{\mu \in \mathbf{S}^1} F(\mu)$$

- Fréchet means might be non-unique, i.e. multiple minimizers
- goal: confidence set C for M: $\mathbf{P}\left(C \supseteq M\right) \geq 1 - \alpha$
- asymptotics in the form of a CLT exist (see talks of Huiling, Stephan and Xavier) but come at the cost of
  - constraints on the distribution of θ,
  - these are difficult to check and
  - in application only finite sample size n available.

Example — Equilateral Triangle



- consider three equidistant point masses with equal weight $\frac{1}{3}$
- $M = \left\{-\frac{2\pi}{3}, 0, \frac{2\pi}{3}\right\}$
- but: for big n empirical mean

$$\underset{\mu \in \mathbf{S}^1}{\arg\min}\, \hat{F}_n(\mu) = \underset{\mu \in \mathbf{S}^1}{\arg\min}\, \frac{1}{n} \sum_{i=1}^{n} d^2(\theta_i, \mu)$$

  is unique with high probability

- asymptotics appear applicable even though they are not!

Stefan Heyder
TU Ilmenau
September 5th 2019

Confidence sets for M

Call $C \subseteq \mathbf{S}^1$ a level $1-\alpha$ confidence set for M if

$$\mathbf{P}\left(C \supseteq M\right) \geq 1-\alpha$$

Desirable properties:

- non-asymptotic: coverage guaranteed for any sample size,
- universal: require only i.i.d. data and
- constructible.

Confidence sets for M

Call $C \subseteq \mathbf{S}^1$ a level $1-\alpha$ confidence set for M if

$$\mathbf{P}\left(C \supseteq M\right) \geq 1-\alpha$$

Desirable properties:

- non-asymptotic: coverage guaranteed for any sample size,
- universal: require only i.i.d. data and
- constructible.

Basic idea: bound deviations of

$$\hat{F}_n(\mu) = \frac{1}{n}\sum_{i=1}^{n} d^2(\theta_i, \mu)$$

from

$$F(\mu) = \mathbf{E}d^2(\theta, \mu)$$

Control deviations

- given α and Δ such that $\mathbf{P}\left(\sup_{\mu \in \mathbf{S}^1} \left|\hat{F}_n(\mu) - F(\mu)\right| \geq \Delta\right) \leq \frac{\alpha}{2}$
- $1 - \frac{\alpha}{2}$ confidence set $C_1 = \left\{\mu \in \mathbf{S}^1 \,:\, \hat{F}_n(\mu) < \inf_{\mu' \in \mathbf{S}^1} \hat{F}_n(\mu') + 2\Delta\right\}$

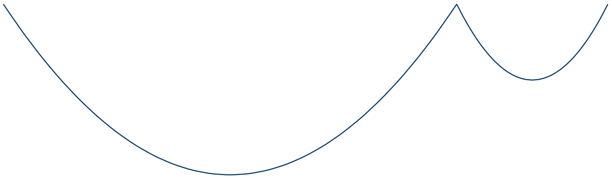


Figure: Band around empirical Fréchet functional

Control deviations

- given α and Δ such that $\mathbf{P}\left(\sup_{\mu\in\mathbf{S}^1}\left|\hat{F}_n(\mu)-F(\mu)\right|\geq\Delta\right)\leq\frac{\alpha}{2}$
- $1-\frac{\alpha}{2}$ confidence set $C_1=\left\{\mu\in\mathbf{S}^1\ :\ \hat{F}_n(\mu)<\inf_{\mu'\in\mathbf{S}^1}\hat{F}_n(\mu')+2\Delta\right\}$

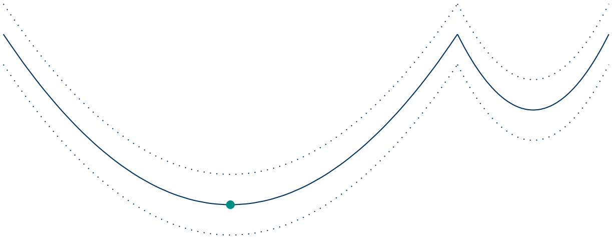


Figure: Band around empirical Fréchet functional

Control deviations

- given α and Δ such that $\mathbf{P}\left(\sup_{\mu \in \mathbf{S}^1} \left|\hat{F}_n(\mu) - F(\mu)\right| \geq \Delta\right) \leq \frac{\alpha}{2}$
- $1 - \frac{\alpha}{2}$ confidence set $C_1 = \left\{\mu \in \mathbf{S}^1 \ : \ \hat{F}_n(\mu) < \inf_{\mu' \in \mathbf{S}^1} \hat{F}_n(\mu') + 2\Delta\right\}$

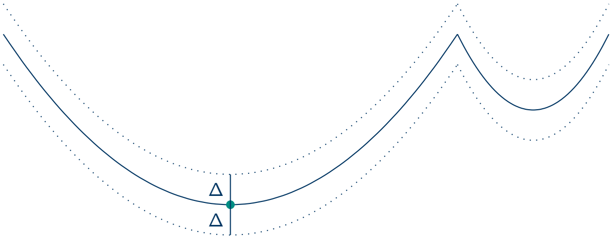


Figure: Band around empirical Fréchet functional

Control deviations

- given α and Δ such that $\mathbf{P}\left(\sup_{\mu\in\mathbf{S}^1}\left|\hat{F}_n(\mu)-F(\mu)\right|\geq\Delta\right)\leq\frac{\alpha}{2}$
- $1-\frac{\alpha}{2}$ confidence set $C_1=\left\{\mu\in\mathbf{S}^1\ :\ \hat{F}_n(\mu)<\inf_{\mu'\in\mathbf{S}^1}\hat{F}_n(\mu')+2\Delta\right\}$

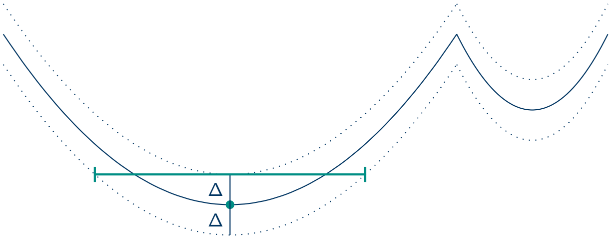


Figure: Band around empirical Fréchet functional

Sketch of construction

Aim:

$$\mathbf{P}\left(\sup_{\mu\in\mathbf{S}^1}\left|\hat{F}_n(\mu)-F(\mu)\right|\geq\Delta\right)\leq\frac{\alpha}{2}$$

Sketch of construction

$$\mathbf{P}\left(\sup_{\mu \in G}\left|\hat{F}_n(\mu) - F(\mu)\right| \geq \Delta_1\right) \leq 2\exp\left(-\frac{2n\Delta_1^2}{\pi^4}\right)$$

Figure: Grid G

Recall Hoeffdings inequality for a single point $\mu \in \mathbf{S}^1$ using $\mathbf{E}\hat{F}_n(\mu) = F(\mu)$ and $0 \leq d^2(\theta, \mu) \leq \pi^2$

$$\mathbf{P}\left(\left|\hat{F}_n(\mu) - F(\mu)\right| \geq \Delta_1\right) \leq 2\exp\left(-\frac{2n\Delta_1^2}{\left(\pi^2\right)^2}\right)$$

Stefan Heyder
TU Ilmenau
September 5th 2019

Sketch of construction

$$\mathbf{P}\left(\sup_{\mu\in G}\left|\hat{F}_n(\mu)-F(\mu)\right|\geq\Delta_1+\Delta_2\right)\leq 2\exp\left(-\frac{2n\Delta_1^2}{\pi^4}\right)+2\cdot 2\exp\left(-\frac{Cn\Delta_2^2}{d^2(\mu,\mu')}\right)$$

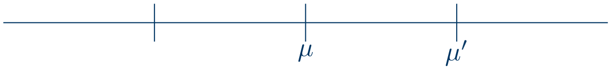


Figure: Grid G

As

$$\left|\hat{F}_n(\mu')-F(\mu')\right|\leq\left|\hat{F}_n(\mu)-F(\mu)\right|+\left|\left(\hat{F}_n(\mu')-F(\mu')\right)-\left(\hat{F}_n(\mu)-F(\mu)\right)\right|,$$

extending to points μ' close to μ „does not cost much“ (Chaining) as $\hat{F}_n-F$ is Lipschitz:

$$\mathbf{P}\left(\left|\left(\hat{F}_n(\mu')-F(\mu')\right)-\left(\hat{F}_n(\mu)-F(\mu)\right)\right|\geq\Delta_2\right)\leq 2\exp\left(-C\frac{n\Delta_2^2}{d^2(\mu,\mu')}\right)$$

Stefan Heyder
TU Ilmenau
September 5th 2019

Sketch of construction

$$\mathbf{P}\left(\sup_{\mu \in G}\left|\hat{F}_n(\mu) - F(\mu)\right| \geq \Delta_1 + \Delta_2 + \Delta_3\right) \leq$$

$$2\exp\left(-\frac{2n\Delta_1^2}{\pi^4}\right) + 4\exp\left(-\frac{Cn\Delta_2^2}{d^2(\mu,\mu')}\right) + 4\cdot 2\exp\left(-\frac{Cn\Delta_3^2}{d^2(\mu',\mu'')}\right)$$

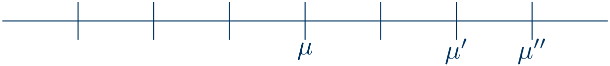


Figure: Grid G

Cost of adding points close to grid G decreases exponentially.

Stefan Heyder
TU Ilmenau
September 5th 2019

Sketch of construction

$$\mathbf{P}\left(\sup_{\mu\in G}\left|\hat{F}_n(\mu)-F(\mu)\right|\geq \Delta_1+\Delta_2+\Delta_3+\Delta_4\right)\leq$$

$$2\exp\left(-\frac{2n\Delta_1^2}{\pi^4}\right)+4\exp\left(-\frac{Cn\Delta_2^2}{d^2(\mu,\mu')}\right)+8\exp\left(-\frac{Cn\Delta_3^2}{d^2(\mu',\mu'')}\right)+8\cdot 2\exp\left(-\frac{Cn\Delta_4^2}{d^2(\mu'',\mu''')}\right)$$

Figure: Grid G

Cost of adding points close to grid G decreases exponentially.

Stefan Heyder
TU Ilmenau
September 5th 2019

Sketch of construction

$$\mathbf{P}\left(\sup_{\mu\in G}\left|\hat{F}_n(\mu)-F(\mu)\right| \geq \Delta_1+\Delta_2+\Delta_3+\Delta_4+\cdots\right) \leq$$

$$2\exp\left(-\frac{2n\Delta_1^2}{\pi^4}\right)+4\exp\left(-\frac{Cn\Delta_2^2}{d^2(\mu,\mu')}\right)+8\exp\left(-\frac{Cn\Delta_3^2}{d^2(\mu',\mu'')}\right)+16\exp\left(-\frac{Cn\Delta_4^2}{d^2(\mu'',\mu''')}\right)+\cdots$$

Figure: Grid G

Cost of adding points close to grid G decreases exponentially.

Stefan Heyder
TU Ilmenau
September 5th 2019

Sketch of construction

$$\mathbf{P}\left(\sup_{\mu \in \mathbf{S}^1} \left|\hat{F}_n(\mu) - F(\mu)\right| \geq \Delta_1 + \Delta_2 + \Delta_3 + \Delta_4 + \cdots + \frac{L}{2\,|G|}\right) \leq$$

$$2\exp\left(-\frac{2n\Delta_1^2}{\pi^4}\right) + 4\exp\left(-\frac{Cn\Delta_2^2}{d^2(\mu,\mu')}\right) + 8\exp\left(-\frac{Cn\Delta_3^2}{d^2(\mu',\mu'')}\right) + 16\exp\left(-\frac{Cn\Delta_4^2}{d^2(\mu'',\mu''')}\right) + \cdots$$

Figure: Grid G

Cover remainder of $\mathbf{S}^1$ using L - Lipschitz continuity of $\hat{F}_n - F$ and solve right hand side for $\frac{\alpha}{2}$

Stefan Heyder
TU Ilmenau
September 5th 2019

TECHNISCHE UNIVERSITÄT ILMENAU

Bounding the derivative

- problem: $\hat{F}_n$ behaves like a quadratic function near global minimizers $\rightsquigarrow$ large confidence sets
- solution: also take derivative $\hat{F}_n'$ into account
- choose ε-cover of $\mathbf{S}^1$, i.e. N_ε balls of radius $\varepsilon > 0$ with centers $\mu_1, \ldots, \mu_{N_\varepsilon}$
- for $\mu \in M$ $\mathbf{E}\hat{F}_n'(\mu) = F'(\mu) = 0$ holds, thus $\mathrm{Var}(\hat{F}_n'(\mu)) = F(\mu) \leq \min_{\mu \in \mathbf{S}^1} \hat{F}_n(\mu) + \Delta =: \sigma^2$ with probability $1 - \frac{\alpha}{2}$
  $\rightsquigarrow$ use Bennett-Hoeffding inequality:

$$\mathbf{P}\left(\exists \mu \in M : \left|\hat{F}_n'(\mu)\right| > \delta\right) \leq \sum_{j=1}^{N_\varepsilon} \mathbf{P}\left(\left|\hat{F}_n'(\mu_j)\right| > \delta\right)$$

- problem: Chaining does not work as $\hat{F}_n'$ may be discontinuous

Bounding the derivative

- problem: $\hat{F}_n$ behaves like a quadratic function near global minimizers $\rightsquigarrow$ large confidence sets
- solution: also take derivative $\hat{F}_n'$ into account
- choose ε-cover of $\mathbf{S}^1$, i.e. N_ε balls of radius $\varepsilon > 0$ with centers $\mu_1, \ldots, \mu_{N_\varepsilon}$
- for $\mu \in M$ $\mathbf{E}\hat{F}_n'(\mu) = F'(\mu) = 0$ holds, thus $\mathrm{Var}(\hat{F}_n'(\mu)) = F(\mu) \leq \min_{\mu \in \mathbf{S}^1} \hat{F}_n(\mu) + \Delta =: \sigma^2$ with probability $1 - \frac{\alpha}{2}$
  $\rightsquigarrow$ use Bennett-Hoeffding inequality:

$$\mathbf{P}\left(\exists \mu \in M : \left|\hat{F}_n'(\mu)\right| > \delta\right) \leq 2N_\varepsilon \exp\left(-\frac{n\sigma^2}{(2\pi)^2}\phi\left(\frac{2\pi\delta}{\sigma^2}\right)\right)$$

- problem: Chaining does not work as $\hat{F}_n'$ may be discontinuous

Putting everything together

- construct $1 - \frac{\alpha}{2}$ confidence set using chaining

$$C_1 = \left\{ \mu \in \mathbf{S}^1 \middle| \hat{F}_n(\mu) \leq \min_{\mu' \in \mathbf{S}^1} \hat{F}_n(\mu') + 2\Delta \right\}$$

- construct $1 - \frac{\alpha}{2}$ confidence set using derivative:

$$C_2 = \left\{ \mu \in \mathbf{S}^1 \middle| \exists j : d(\mu, \mu_j) \leq \varepsilon \wedge \inf_{\mu \in \mathbf{S}^1 : d(\mu, \mu_j)} \left| \hat{F}_n'(\mu) \right| \leq \delta \right\}$$

- intersection $C_1 \cap C_2$ is a $1 - \alpha$ confidence set
- iterate covering only $C_1 \cap C_2$

Stefan Heyder
TU Ilmenau

Simulation 1: equilateral triangle

$$\mathbf{P}(\theta = -\tfrac{2}{3}\pi) = \mathbf{P}(\theta = 0) = \mathbf{P}(\theta = \tfrac{2}{3}\pi) = \tfrac{1}{3}$$

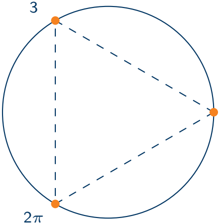


Figure: one simulation for $n = 10^4$: $C_1 \cap C_2 \supseteq M$

n	$\boldsymbol{\lambda}(C_1 \cap C_2)$	Coverage %
10^2	6.21 ($\pm$0.13)	100
10^3	1.93 ($\pm$0.01)	100
10^4	0.56 ($\pm$0.00)	100
10^5	0.18 ($\pm$0.00)	100

- 1000 repetitions per sample size n
- $1 - \alpha = 90\%$
- $\boldsymbol{\lambda}(C_1 \cap C_2)$: Lebesgue measure

Simulation 1: equilateral triangle

$$\mathbf{P}(\theta = -\tfrac{2}{3}\pi) = \mathbf{P}(\theta = 0) = \mathbf{P}(\theta = \tfrac{2}{3}\pi) = \tfrac{1}{3}$$

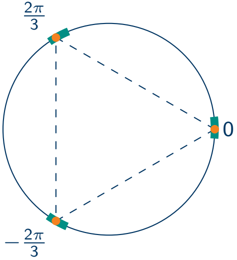


Figure: one simulation for $n = 10^4$: $C_1 \cap C_2 \supseteq M$

n	$\lambda(C_1 \cap C_2)$	Coverage %
10^2	6.21 ($\pm$0.13)	100
10^3	1.93 ($\pm$0.01)	100
10^4	0.56 ($\pm$0.00)	100
10^5	0.18 ($\pm$0.00)	100

- 1000 repetitions per sample size n
- $1 - \alpha = 90\%$
- $\lambda(C_1 \cap C_2)$: Lebesgue measure

Simulation 2: partly uniform density

$$\theta \sim 0.6\,\delta_0 + \tfrac{1}{2\pi}\,\mathbf{1}_{(-\pi,-0.6\pi]\cup[0.6\pi,\pi]}\,\boldsymbol{\lambda}$$

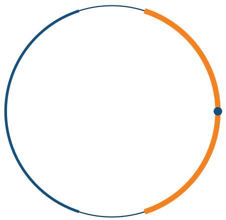


Figure: one simulation for $n = 10^4$: $C_1 \cap C_2 \supseteq M$

n	$\boldsymbol{\lambda}\big((C_1 \cap C_2) \setminus M\big)$	Coverage %
10^2	2.44 ($\pm$0.32)	100
10^3	0.62 ($\pm$0.11)	100
10^4	0.18 ($\pm$0.03)	100

- $M = [-0.4\pi, 0.4\pi]$
- 100 repetitions per sample size n
- $1 - \alpha = 90\%$
- $\boldsymbol{\lambda}\big((C_1 \cap C_2) \setminus M\big)$: excess length

Simulation 2: partly uniform density

$$\theta \sim 0.6\,\delta_0 + \frac{1}{2\pi}\,\mathbf{1}_{(-\pi,-0.6\pi]\cup[0.6\pi,\pi]}\,\boldsymbol{\lambda}$$

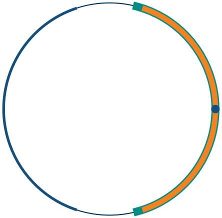


Figure: one simulation for $n = 10^4$: $C_1 \cap C_2 \supseteq M$

n	$\boldsymbol{\lambda}\big((C_1 \cap C_2) \setminus M\big)$	Coverage %
10^2	2.44 ($\pm$0.32)	100
10^3	0.62 ($\pm$0.11)	100
10^4	0.18 ($\pm$0.03)	100

- $M = [-0.4\pi, 0.4\pi]$
- 100 repetitions per sample size n
- $1 - \alpha = 90\%$
- $\boldsymbol{\lambda}\big((C_1 \cap C_2) \setminus M\big)$: excess length

Summary

constructed non-asymptotic, universal for Fréchet means on the circle

- no distributional assumptions
- automatically adapts to smeariness
- guaranteed coverage for every sample size
- excess length typically $\mathcal{O}\left(\sqrt{\log n / n}\right)$
- can be generalized to compact Riemannian manifolds

future research

- can one extend chaining to derivative?

Thank you for your attention!

[1] M. Glock and T. Hotz. Constructing Universal, Non-asymptotic Confidence Sets for Intrinsic Means on the Circle. In *International Conference on Geometric Science of Information*, pages 477–485. Springer, 2017.

[2] T. Hotz and S. Huckemann. Intrinsic means on the circle: uniqueness, locus and asymptotics. *Annals of the Institute of Statistical Mathematics*, 67(1):177–193, Feb. 2015.